

Szybki rozwój nauki i technologii jest przyczyną niewiarygodnego wzrostu ilości i złożoności produkowanych danych, obserwowanego w różnych obszarach, takich jak informatyka, medycyna, czy nauki podstawowe, a zwłaszcza fizyka cząstek elementarnych i astronomia. Obecnie pojedyncze zbiory danych mogą zawierać nawet petabajty informacji, stąd też termin *big data*, który pojawił się w fachowej literaturze w ostatnich latach. Powstanie *big data* tworzy nowe wyzwania związane z przetwarzaniem ogromnych ilości danych, takie jak monitorowanie ich jakości oraz ich wizualizacja.

Jednym z przykładów systemów informatycznych, wewnątrz których generowane są olbrzymie ilości danych, jest system działający w międzynarodowym instytucie badawczym stworzonym w celu analizy zjawisk fizycznych towarzyszących zderzeniom cząstek elementarnych - CERN. Kluczowym narzędziem pracy wykonywanej w CERN jest Wielki Zderzacz Hadronów (ang. Large Hadron Collider, LHC), który jest jednocześnie jednym z największych projektów naukowych realizowanych obecnie na świecie. LHC najbardziej znany jest z eksperymentu potwierdzającego istnienie słynnego bozonu Higgsa. Doprowadził on do przyznania Nagrody Nobla dwóm fizykom, którzy jeszcze w latach 60. ubiegłego wieku postulowali istnienie tej cząstki: François Englertowi oraz Peterowi Higgsowi.

Oprócz wspomnianego eksperymentu potwierdzającego istnienie bozonu Higgsa, na LHC odbywają się również inne eksperymenty. Trzecim pod względem wielkości eksperymentem badawczym jest ALICE (ang. A Large Ion Collider Experiment). Członkiem tego eksperymentu jest także grupa badaczy z Politechniki Warszawskiej, z którymi przewidywana jest współpraca w ramach proponowanego projektu. Eksperyment ALICE dedykowany jest w szczególności zderzeniom ciężkich jonów, w wyniku których powstaje stan materii podobny do tego, jaki zaistniał po Wielkim Wybuchu.

Proponowany projekt realizowany będzie właśnie we współpracy z CERN, ponieważ eksperymenty działające na LHC produkują olbrzymie ilości danych, które następnie muszą być przygotowane i przeanalizowane. W przypadku eksperymentu ALICE od 2009 roku zebrano ponad 24 PB danych, które są gotowe do analiz fizycznych. Środowisko to jest zatem modelowym przykładem występowania tzw. *big data* gdzie generowane są zestawy danych o dużej objętości i złożoności. Dane, które zbiera eksperyment ALICE zawierają setki tysięcy parametrów pochodzących z ogromnej liczby sensorów każdego z 18 detektorów wchodzących w skład eksperymentu. W tym miejscu należy podkreślić, że obecne systemy monitoringu jakości danych używane w ALICE oparte są na manualnych porównywaniach histogramów rejestrowanych zjawisk, a ich wizualizacja wykorzystuje tylko mały procent danych w celu wyświetlenia np. śladów cząstek w detektorach.

Celem niniejszego projektu jest rozszerzenie możliwości systemu monitoringu jakości danych o metody uczenia maszynowego oraz stworzenie nowych systemów wizualizacji zbieranych danych. Metody sztucznej inteligencji pozwolą na odciążenie aktualnie istniejącej infrastruktury analizy jakości danych, opartej w dużym stopniu na pracy ludzi, poprzez uwzględnienie historycznych danych oraz czynników zewnętrznych (np. temperatury otoczenia). W ramach projektu planuje się rozszerzyć metody manualnej analizy danych o algorytmy sztucznej inteligencji, takie jak sieci neuronowe czy algorytmy boostingu, a także opracować nowe, bardziej wydajne metody uczenia maszynowego służące do kontroli jakości ogromnych ilości danych w trybie rzeczywistym.

Nowe systemy wizualizacji pozwolą natomiast na bardziej efektywną prezentację zarówno efektów fizycznych, jak i właściwości samego detektora. Zaproponowane metody wizualizacji oparte będą o urządzenia wirtualnej rzeczywistości, takie jak gogle 3D oraz wyświetlacze 3D. Zastosowanie takiego podejścia wykracza poza istniejące metody prezentacji, które wywodzą się z pokazywania danych 3D na płaskich wyświetlaczach i stosują daleko idące uproszczenia struktury danych. Stworzenie systemu wizualizacji dużych ilości danych może być również pomocne przy ich przeszukiwaniu i analizowaniu reprezentowanych przez nie zjawisk.

Warto również podkreślić, że trójwymiarowe wizualizacje zderzeń cząstek są bardzo przydatne w celach popularyzatorskich. Na ich podstawie można w przystępny sposób wytłumaczyć szerokiemu społeczeństwu skomplikowane efekty fizyczne, które badane są przy pomocy tak drogich urządzeń jak LHC. Tego typu narzędzia można również wykorzystać w szkołach na lekcjach fizyki, do pokazywania niektórych zjawisk fizycznych (np. rozpadów cząstek nietrwałych).